SUPPLEMENTARY INFORMATION

Title: Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*

Short title: *Burkholderia ambifaria*: Genomics and biology of a biopesticide

Authors:
Alex J. Mullins[1], James. A. H. Murray[1], Matthew J. Bull[1], Matthew Jenner[2], Cerith Jones[1,6], Gordon Webster[1], Angharad E. Green[3], Daniel R. Neill[3], Thomas R. Connor[1], Julian Parkhill[4], Gregory L. Challis[2,5] and Eshwar Mahenthiralingam[1]

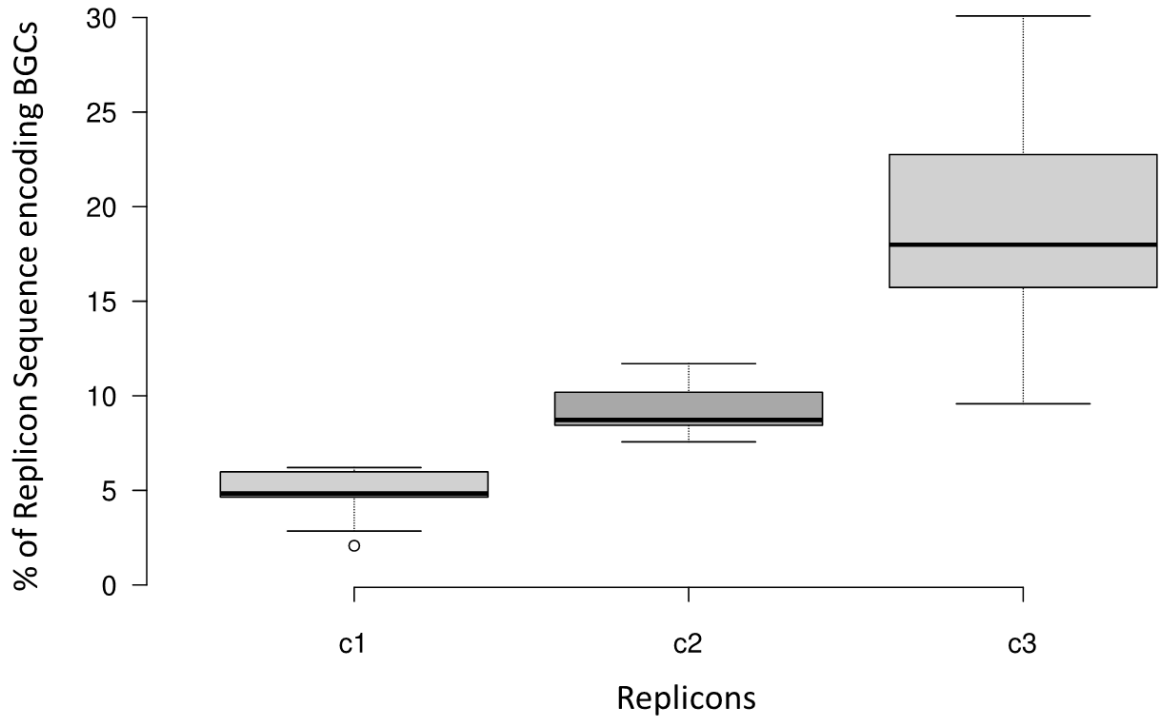
Affiliations:
[1]Microbiomes, Microbes and Informatics Group, Organisms and Environment Division, School of Biosciences, Cardiff University, Cardiff CF10 3AX, United Kingdom
[2]Department of Chemistry and Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry CV4 7AL, United Kingdom
[3]Institute of Infection and Global Health, University of Liverpool, Liverpool L69 7BE, United Kingdom
[4]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom
[5]Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia
[6]Current address: Faculty of Computing, Engineering and Science, University of South Wales, Pontypridd, CF37 1DL, United Kingdom

Correspondence:

Submission and handling: Prof. Eshwar Mahenthiralingam, Microbiomes, Microbes and Informatics Group, Organisms and Environment Division, School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff CF10 3AX, United Kingdom; Tel. +44 (0)29 20875875; Fax. +44 (0)29 20874305;
Email: MahenthiralingamE@cardiff.ac.uk (ORCID: 0000-0001-9014-3790)

Joint correspondence: Alex J. Mullins, Microbiomes, Microbes and Informatics Group, Organisms and Environment Division, School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff CF10 3AX, United Kingdom; Tel. +44 (0)29 20874648
Email: MullinsA@cardiff.ac.uk (ORCID: 0000-0001-5804-9008)

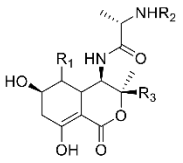
Table of Contents



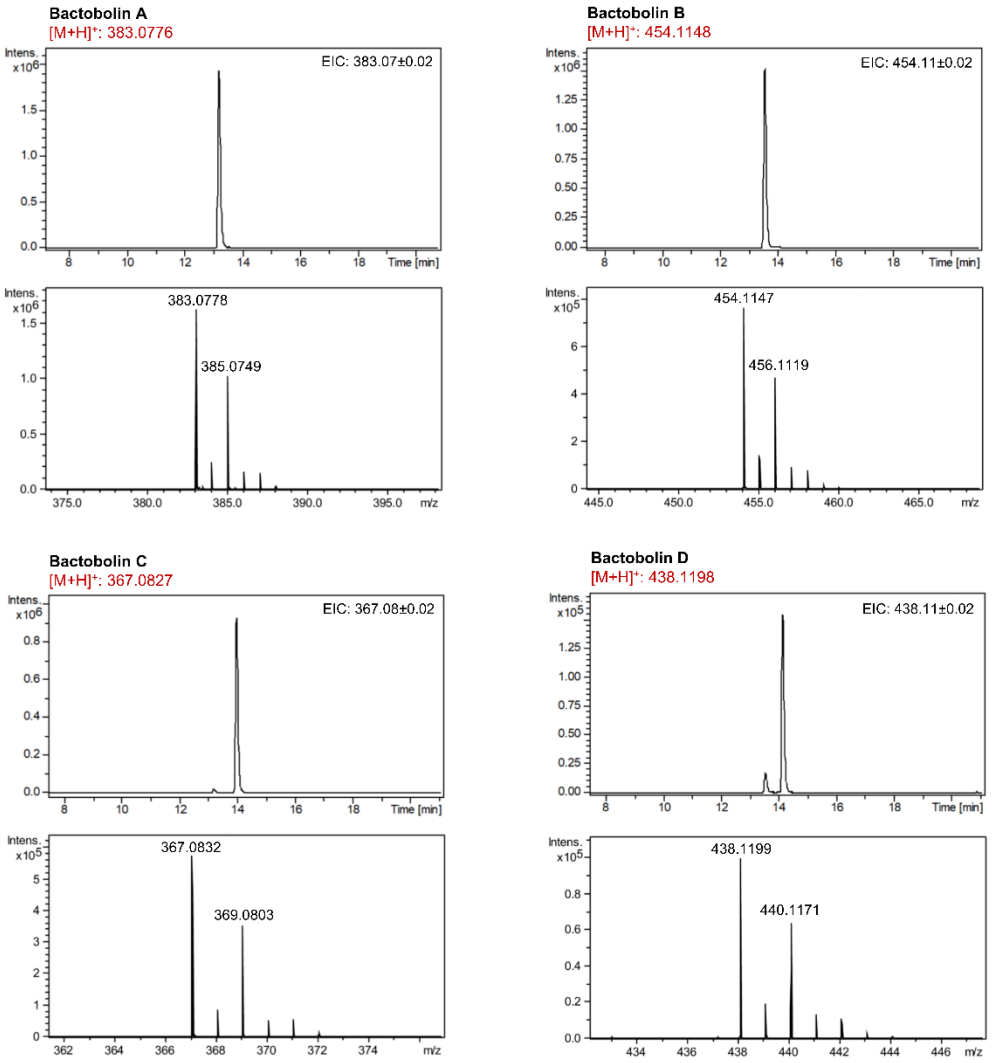
Supplementary Figure 1. Distribution of secondary metabolite biosynthetic potential across the three replicons of 64 *B. ambifaria* strains. Centre lines represent the median; box limits indicate the 25th and 75th percentiles; and whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. Only 63 strains are represented by the box plot of replicon c3 due to the lack of a third replicon in *B. ambifaria* BCC1105. Boxplots were generated using the web tool BoxPlotR[5].

Supplementary Figure 2. LC-MS data for observed and theoretical mass(es), and mass spectrum for each antimicrobial metabolite detected in screened *B. ambifaria* strains. (a) Bactobolins; **(b)** Pyrrolnitrin; **(c)** Enacyloxin IIa; **(d)** Hydroxyquinolines; and **(e)** Burkholdines. To profile the metabolites produced by each strain under agar-growth condition (see Methods), *n* = 1 LC-MS analysis per strain (10 strains examined) (Table 1).

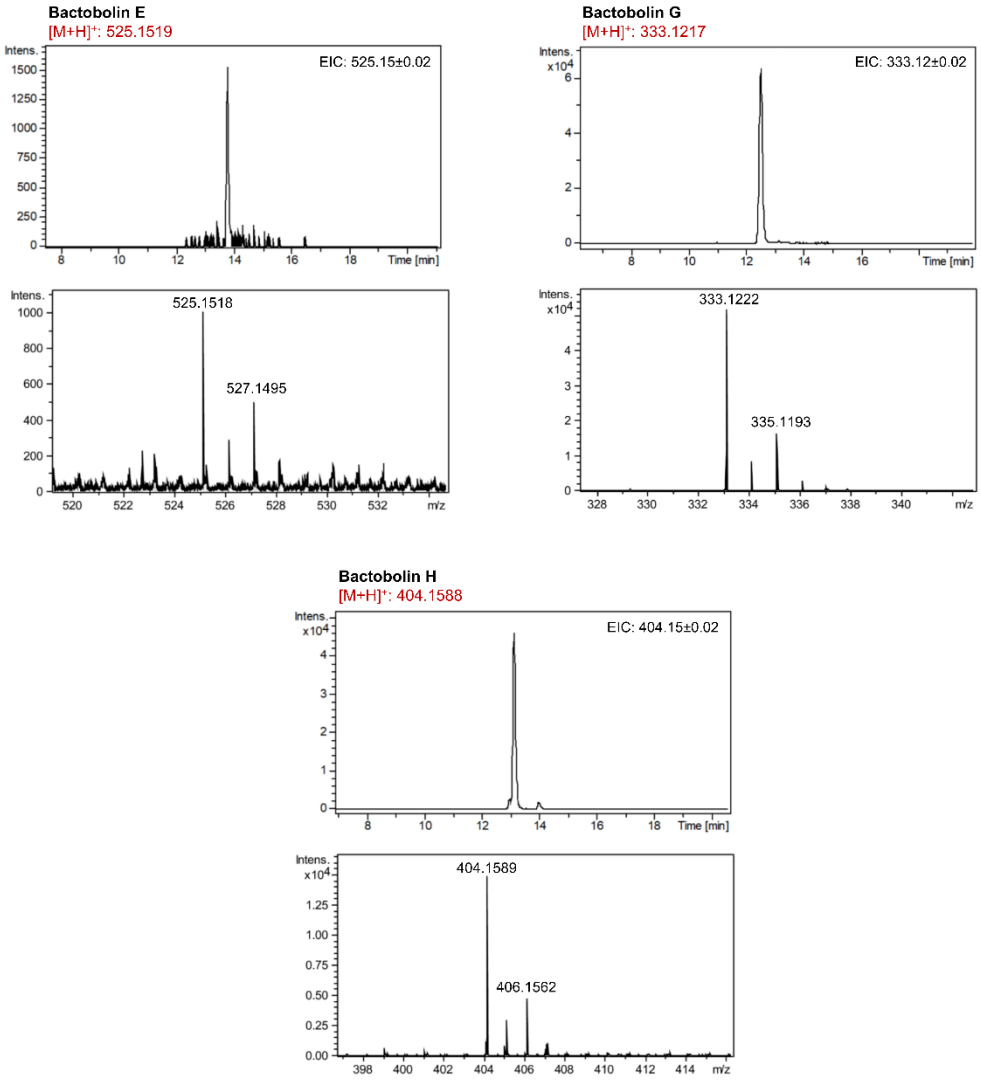
Supplementary Figure 2 (a) Bactobolins (A to D)



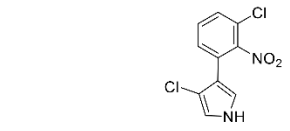
Bactobolins	Chemical Formula	Exact Mass
A: R_1=OH, R_2=H, R_3=$CHCl_2$	$C_{14}H_{20}Cl_2N_2O_6$	382.0698
B: R_1=OH, R_2=L-Ala, R_3=$CHCl_2$	$C_{17}H_{25}Cl_2N_3O_7$	453.1070
C: R_1=H, R_2=H, R_3=$CHCl_2$	$C_{14}H_{20}Cl_2N_2O_5$	366.0749
D: R_1=H, R_2=L-Ala, R_3=$CHCl_2$	$C_{17}H_{25}Cl_2N_3O_6$	437.1120
E: R_1=OH, R_2=L-Ala-L-Ala, R_3=$CHCl_2$	$C_{20}H_{30}Cl_2N_4O_8$	524.1441
G: R_1=H, R_2=H, R_3=CH_2Cl	$C_{14}H_{21}ClN_2O_5$	332.1139
H: R_1=H, R_2=L-Ala, R_3=CH_2Cl	$C_{17}H_{26}ClN_3O_6$	403.1510



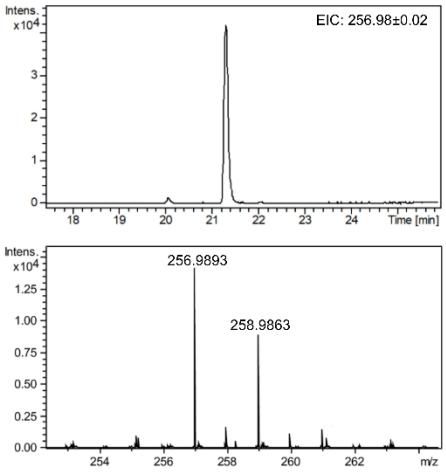
Supplementary Figure 2 (a) Bactobolins (E to H)



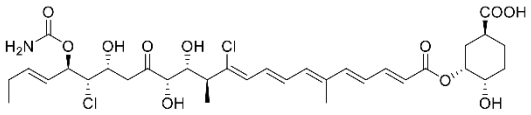
Supplementary Figure 2 (b) Pyrrolnitrin



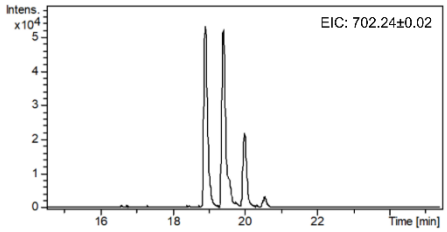
Pyrrolnitrin
Chemical Formula: $C_{10}H_6Cl_2N_2O_2$
Exact Mass: 255.9806
$[M+H]^+$: 256.9884

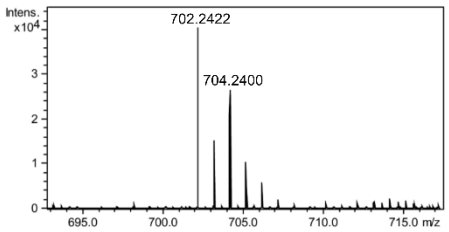


Supplementary Figure 2 (c) Enacyloxin IIa

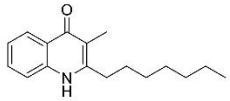


Enacyloxin IIa
Chemical Formula: $C_{33}H_{45}Cl_2NO_{11}$
Exact Mass: 701.2370
$[M+H]^+$: 702.2447

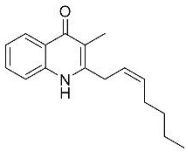




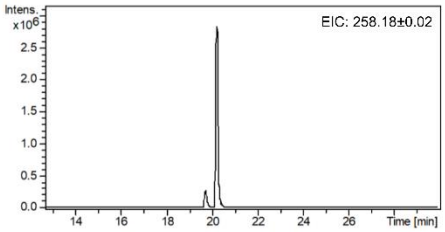
Supplementary Figure 2 (d) Hydroxyquinolines

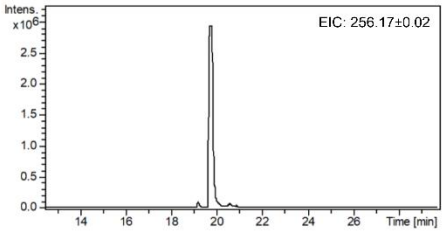


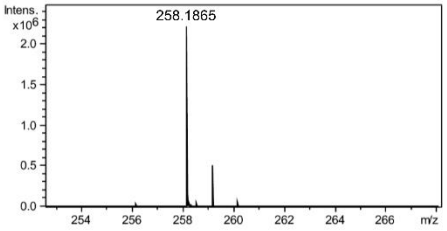
2-heptyl-3-methylquinolin-4(1*H*)-one
Chemical Formula: $C_{17}H_{23}NO$
Exact Mass: 257.1780
$[M+H]^+$: 258.1858

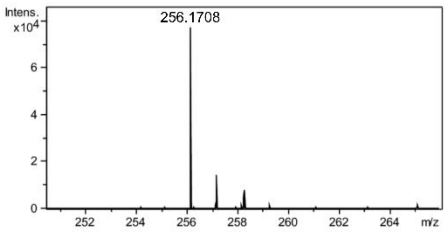


(*Z*)-2-(hept-2-en-1-yl)-3-methylquinolin-4(1*H*)-one
Chemical Formula: $C_{17}H_{21}NO$
Exact Mass: 255.1623
$[M+H]^+$: 256.1701

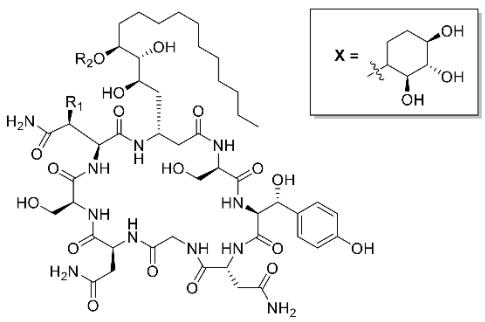


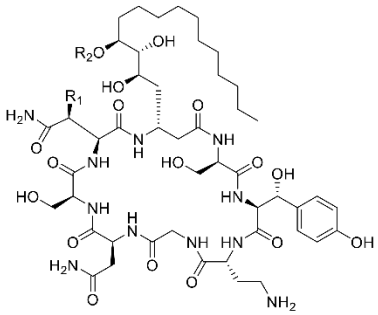






Supplementary Figure 2 (e) Burkholdines

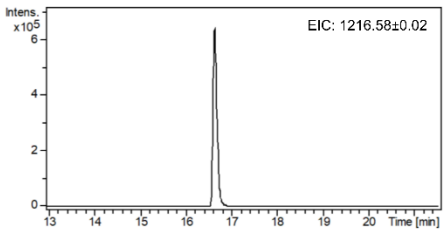


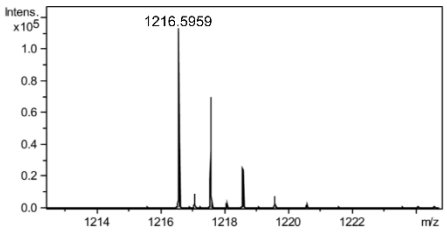


Burkholdine	Chemical Formula	Exact Mass
Bk-1229: R_1=OH, R_2=β-**X**	$C_{52}H_{83}N_{11}O_{23}$	1229.5663
Bk-1097: R_1=OH, R_2=H	$C_{47}H_{75}N_{11}O_{19}$	1097.5240
Bk-1213: R_1=H, R_2=β-**X**	$C_{52}H_{83}N_{11}O_{22}$	1213.5714

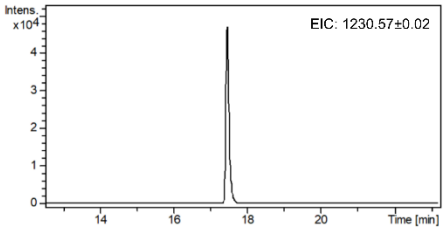
Burkholdine	Chemical Formula	Exact Mass
Bk-1215: R_1=OH, R_2=α-**X**	$C_{52}H_{85}N_{11}O_{22}$	1215.5870
Bk-1199: R_1=H, R_2=β-**X**	$C_{52}H_{85}N_{11}O_{21}$	1199.5921

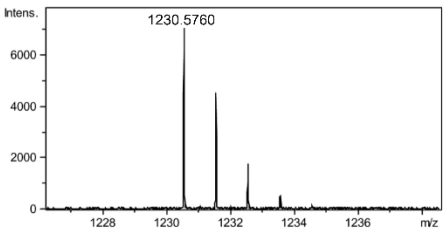
Bk-1215
[M+H]$^+$: 1216.5948



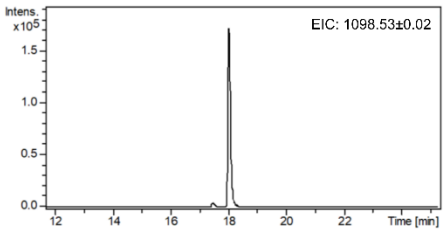


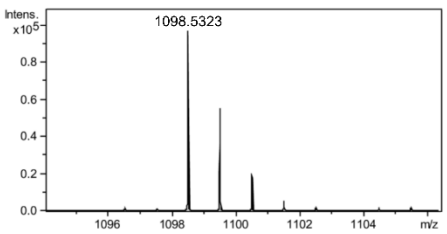
Bk-1229
[M+H]$^+$: 1230.5741



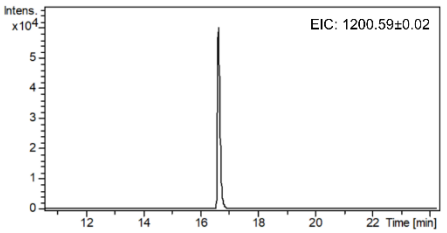


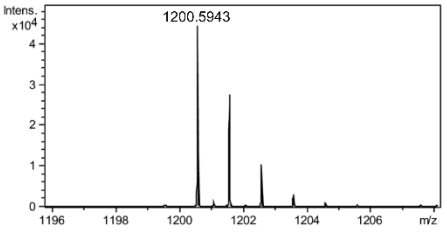
Bk-1097
[M+H]$^+$: 1098.5318

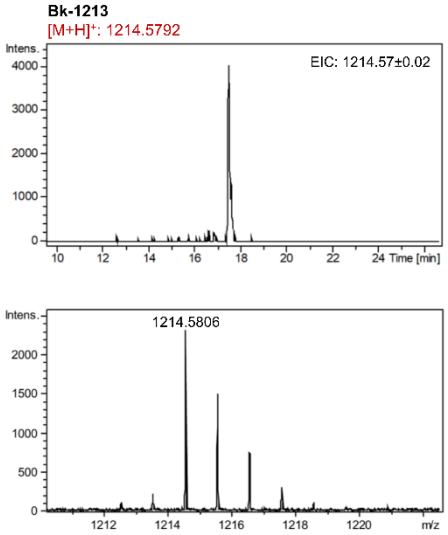


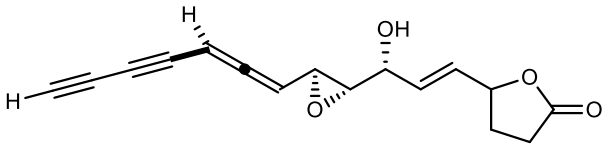


Bk-1199
[M+H]$^+$: 1200.5999

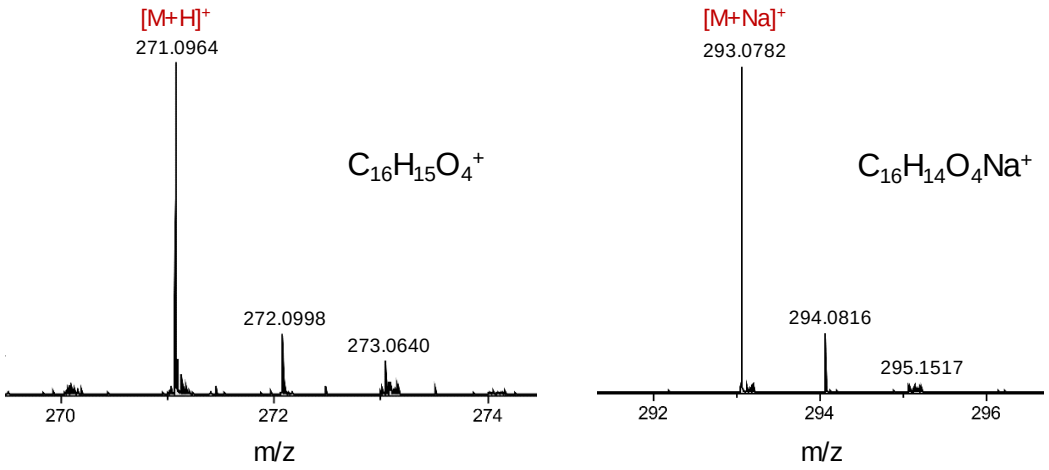




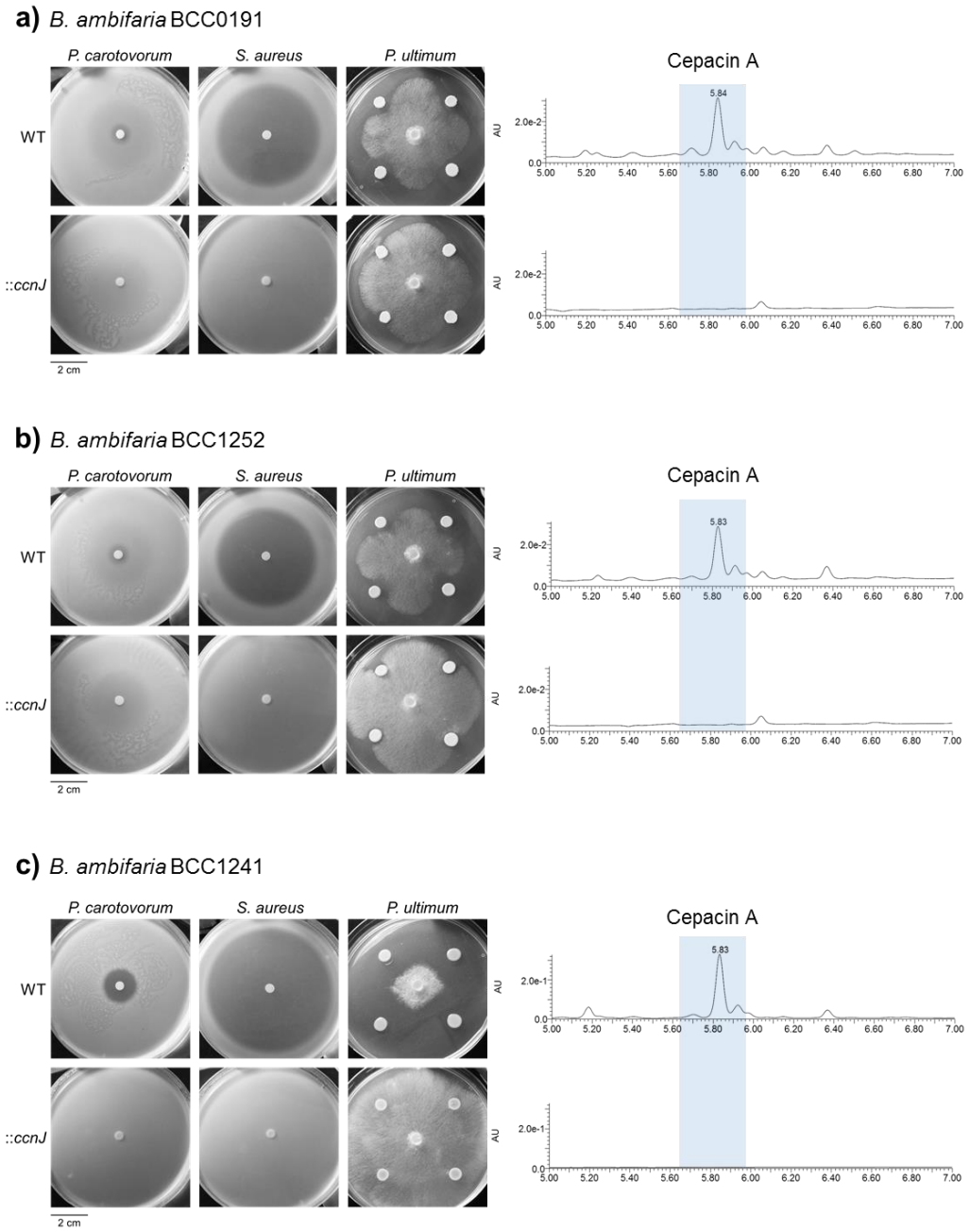

Bk-1213
[M+H]+: 1214.5792
Intens.
4000
3000
2000
1000
0
EIC: 1214.57±0.02
10
12
14
16
18
20
22
24 Time [min]
Intens.
2000
1500
1000
500
0
1214.5806
1212
1214
1216
1218
1220
m/z



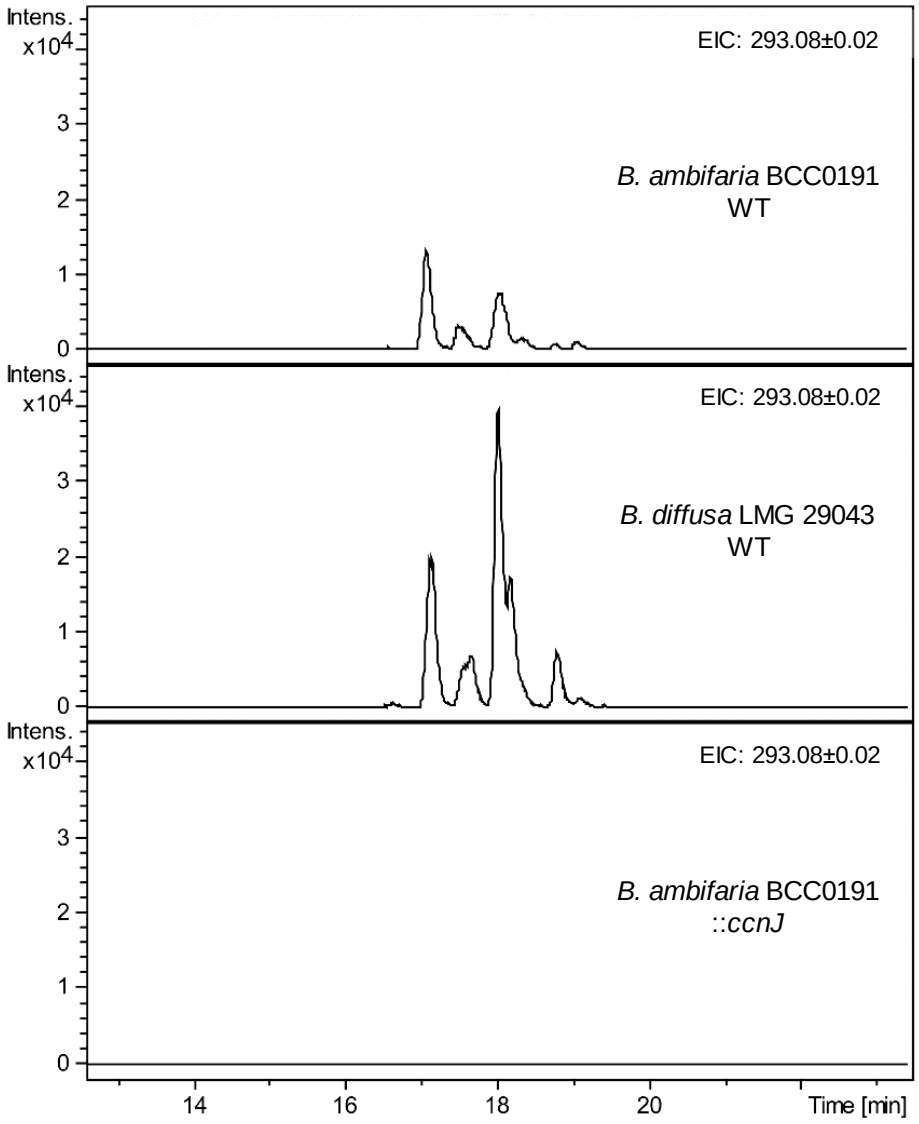
Cepacin A
Chemical Formula: $C_{16}H_{14}O_4$
Exact Mass: 270.0892
$[M+H]^+$: 271.0965
$[M+Na]^+$: 293.0784



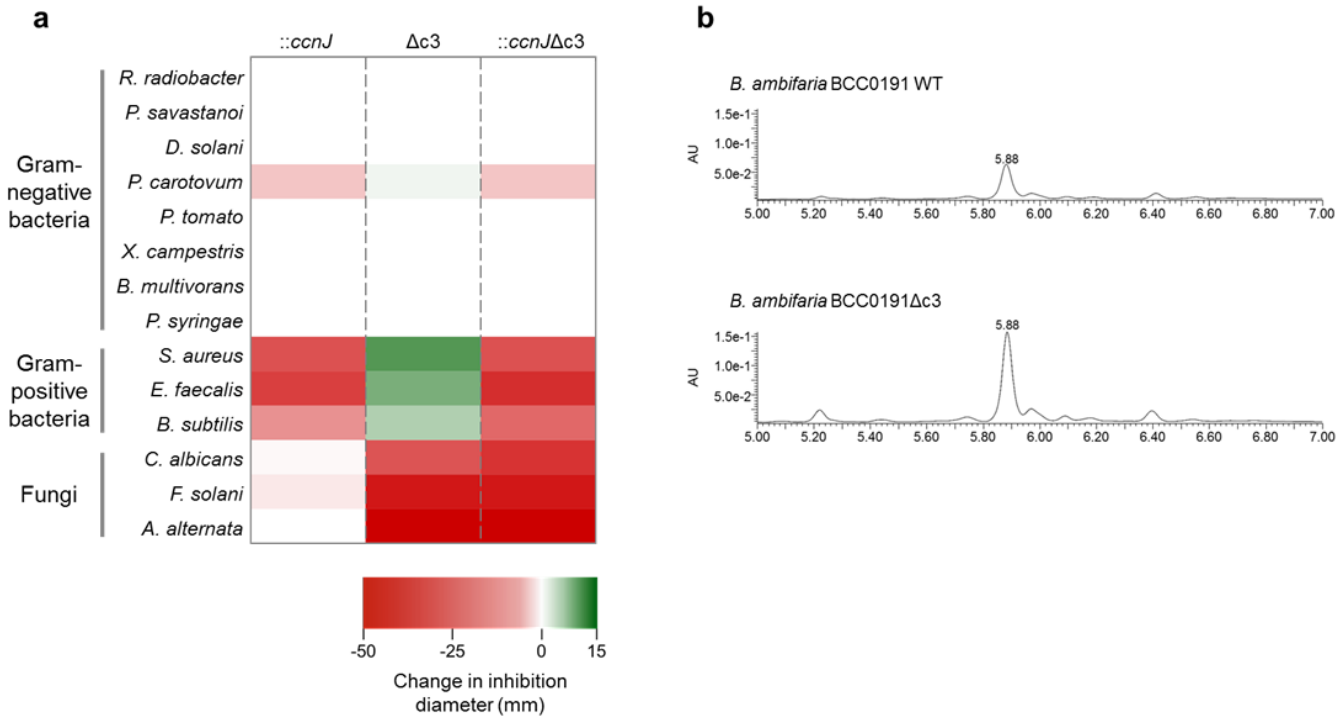
Supplementary Figure 3. High-resolution mass spectrometry analysis of *B. ambifaria* BCC0191 cepacin A. Measured spectra of cepacin A $[M+H]^+$ (left) and $[M+Na]^+$ (right). The generated molecular formulae for each species are shown and are in agreement with the molecular formula of cepacin A. *n* = 3 independent LC-MS analyses on agar grown cultures of strain BCC0191.



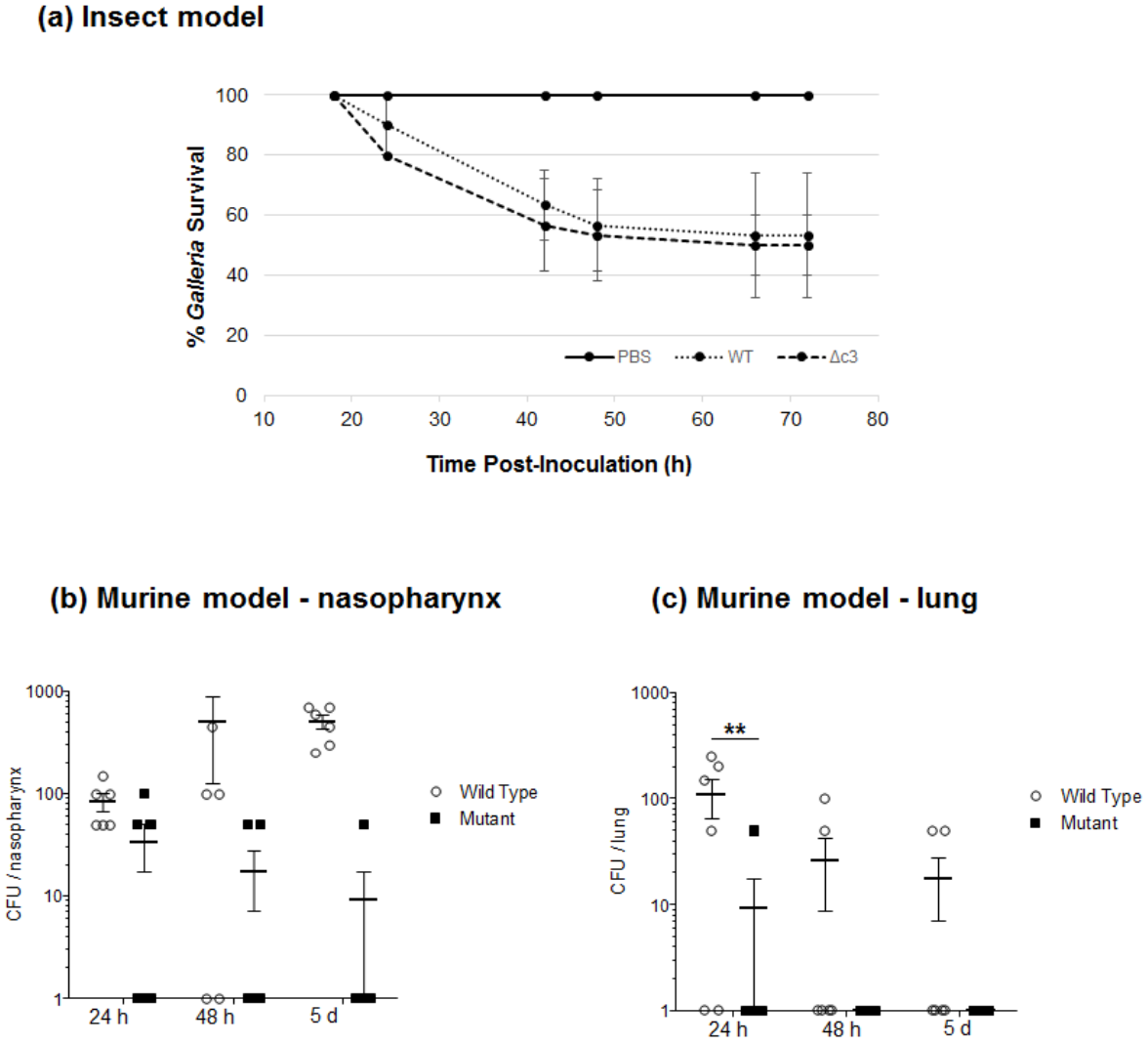
Supplementary Figure 4. Antimicrobial activity, HPLC chromatograms and presence/absence of cepacin-related HPLC peaks for *B. ambifaria* wild-type (WT) and insertional mutants (::*ccnJ*). Antimicrobial activity against *P. carotovorum*, *S. aureus*, and *P. ultimum;* and HPLC chromatograms at 260 nm of the WT (see Supplementary Methods) and cepacin-deficient deriviatives of *B. ambifaria* strains (a) BCC0191, (b) BCC1252, and (c) BCC1241 are shown. The absence of cepacin production in the cepacin-deficient derivatives of *B. ambifaria* strains BCC0477, BCC1259 and BCC1218 was also confirmed by HPLC analysis. *n* = 2 biological replicates per overlay; and *n* = 2 HPLC analyses for each strain wild type and mutant.



Supplementary Figure 5. Comparative cepacin A metabolite analysis of the original producer strain *B. diffusa* LMG 029043 and *B. ambifaria* BCC0191. Extracted ion chromatograms at m/z 293.08 ± 0.02, corresponding to the $[M + Na]^+$ ion of cepacin A, from LC-MS analyses of crude extracts from agar-grown cultures of *B. ambifaria* BCC0191 (top), *B. diffusa* LMG 29043 (middle), and the insertional mutant *B. ambifaria* BCC0191::*ccnJ* (bottom). *n* = 1 independent LC-MS analysis per strain/derivative.



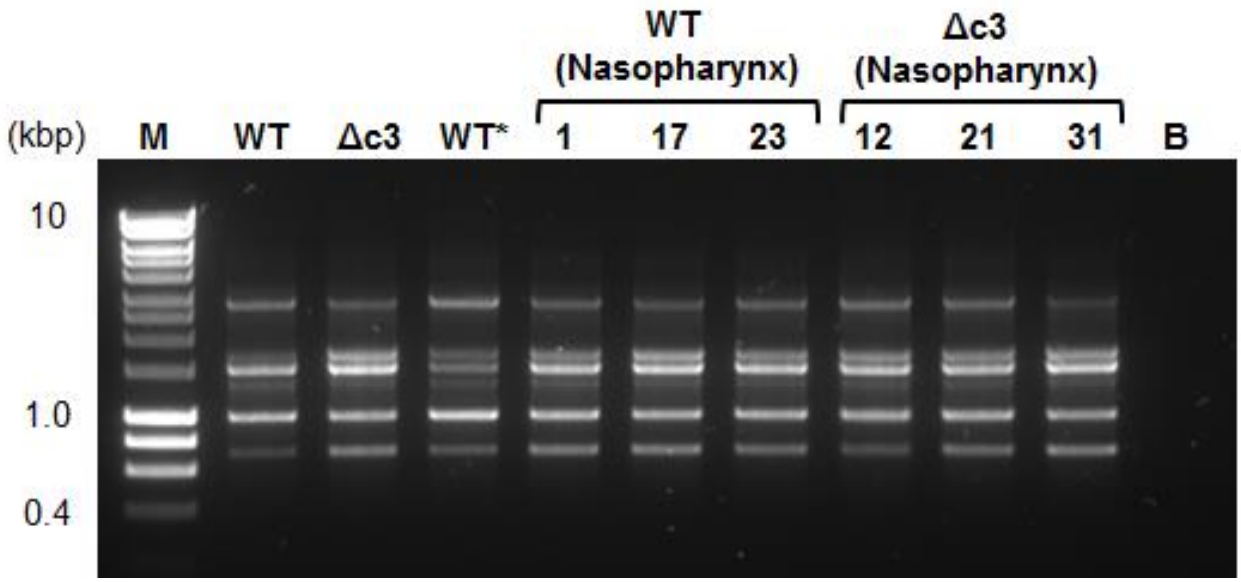
Supplementary Figure 6. Impact of mutations on *B. ambifaria* BCC0191 antimicrobial activity and cepacin production. (a) Three mutants were compared to BCC0191 wild-type (WT): cepacin-deficient derivative (::*ccnJ*), third replicon knockout (Δc3), and a combined mutation (::*ccnJ*Δc3). *n* = 2 independent overlays per condition, and the mean average used to generate the heat map. Scale represents change in zone of inhibition diameter (mm) compared to BCC0191 WT (red = reduced zone, white = no change, green = increased zone). **(b)** HPLC chromatograms at 260 nm of *B. ambifaria* BCC0191 WT and BCC0191Δc3 (*n* = 6 independent HPLC analyses per strain) highlighting the impact of third replicon deletion on cepacin production.



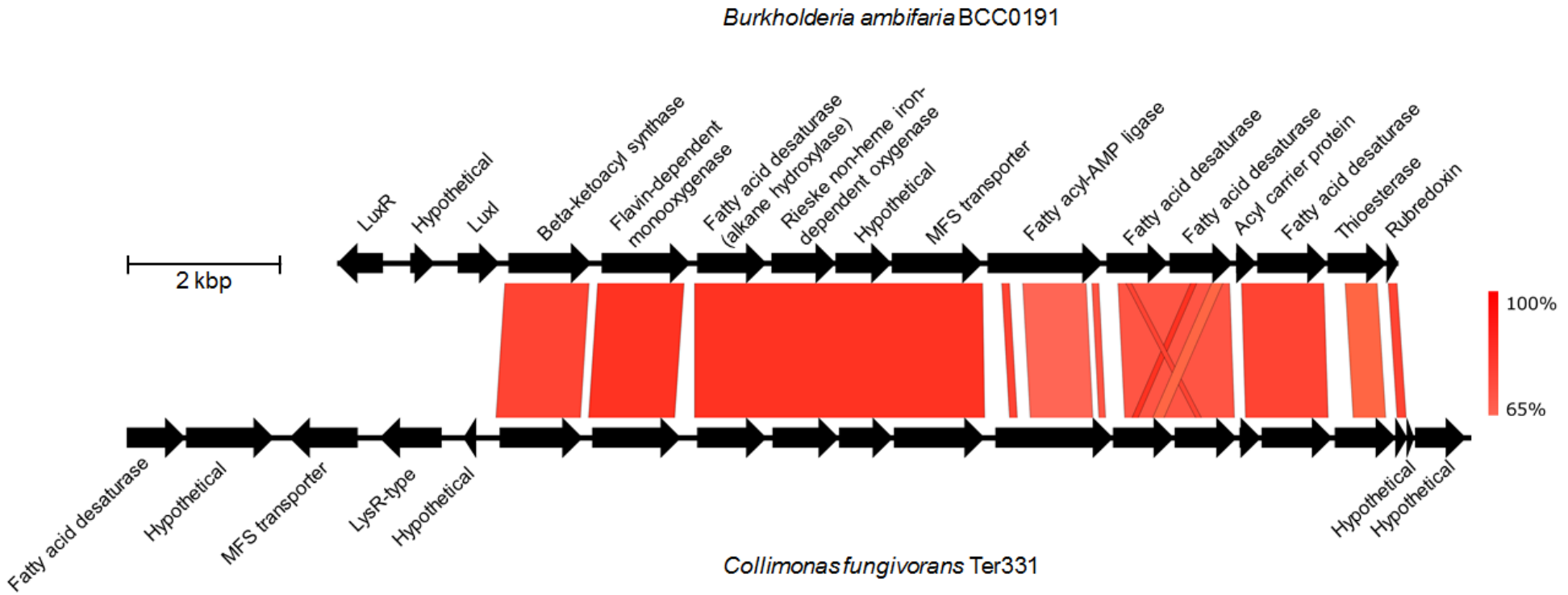
Supplementary Figure 7. Persistence of *B. ambifaria* BCC0191 and its third replicon mutant in models of infection.

***Galleria mellonella* wax moth larvae model (a).** Groups of 10 *Galleria mellonella* wax moth larvae, were injected in the last proleg on the right-side of the abdomen using a restraining technique as described[6] (*n* = 30 larvae per condition). The larvae were incubated at 37°C for 72 hours, and their survival status was monitored periodically. Larvae were recorded as dead when they failed to respond to physical agitation. Centre bar represents the mean, and error bars represent standard deviation.

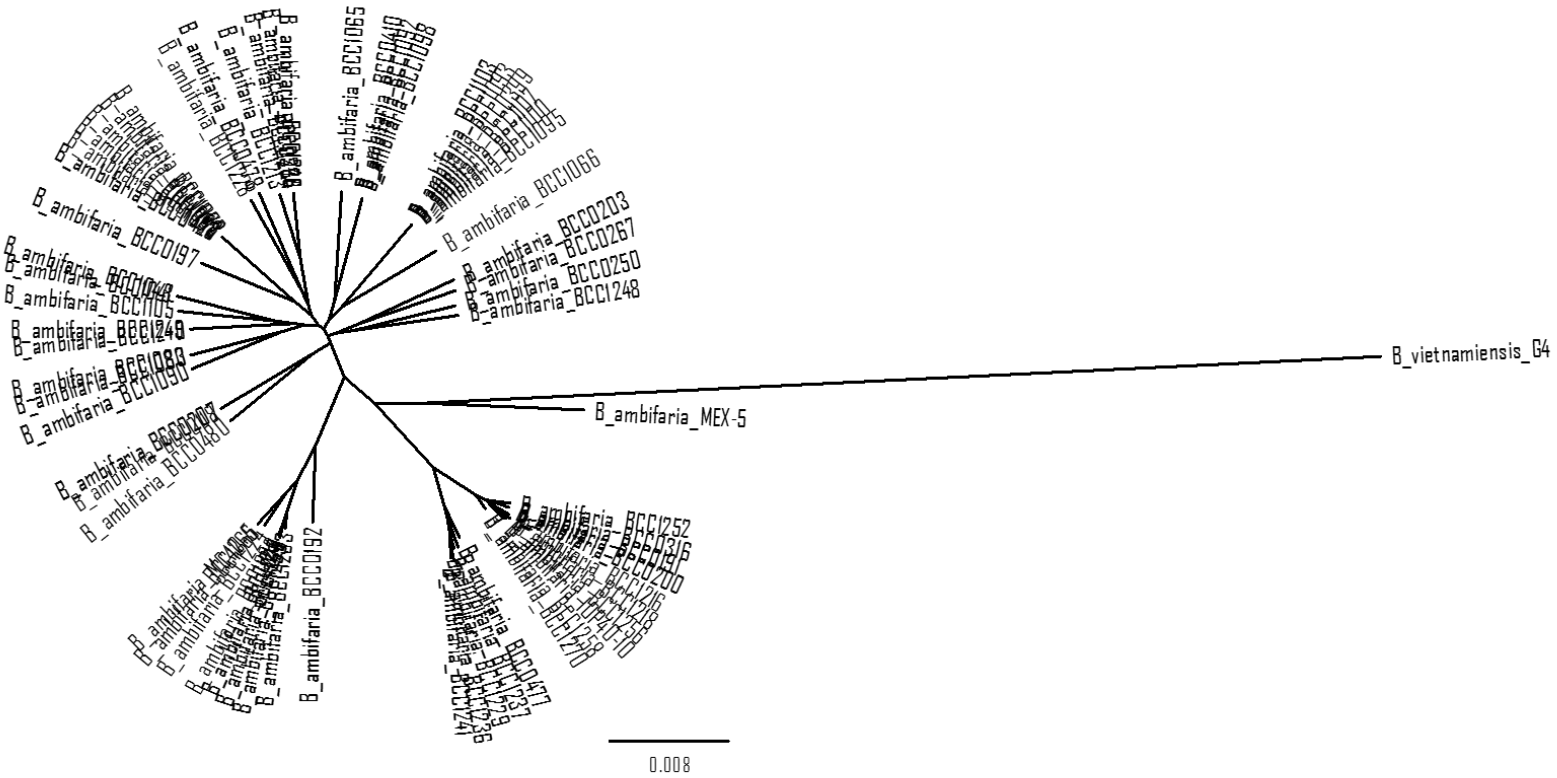
Murine model of respiratory tract infection: (b) nasopharynx and (c) lung. Mice (*n* = 6 mice per time point) were infected via inhalation with a dose of 2 x 10^6 *B. ambifaria* BCC0191 (wild type) and its corresponding third replicon deletion mutant, BCC0191Δc3 (mutant), respectively. The mean number of colony forming units recovered from the nasopharynx (a) and lung (b) are shown after 24 hours, 48 hours and 5 days of infection. The centre bar represents the mean, and error bars represent the standard error. Persistence of the third replicon mutant after 24 hours in the lung was significantly lower than the wild type (** = *P* = 0.0038 in two-way ANOVA with Sidak's multiple comparisons test of wild type vs mutant). No colony forming units of the third replicon mutant were detected in the lungs of mice after 48 hours and 5 days of infection.



Supplementary Figure 8. PCR genotyping of *B. ambifaria* BCC0191 and BCC0191Δc3 recovered from the murine respiratory infection model. DNA was extracted from bacteria recovered from the murine infection model (*n* = 31 samples) using chelex-100 resin and random amplified polymorphic DNA (RAPD) fingerprinting PCR[7] was performed (*n* = 1 independent PCR per sample) to validate their genetic identity. PCR products separated by agarose gel electrophoresis, and lanes are as follows: WT, strain BCC0191; Δc3, BCC0191Δc3; WT*, BCC0191 high quality DNA extracted using Maxwell16 instrument; 1, 17 and 23, BCC0191 WT colonies recovered from the nasopharynx of mice at 24 hours, 48 hours and 5 days of infection, respectively; 12, 21, and 31, BCC0191::Δc3 mutant colonies recovered from the nasopharynx of mice at 24 hours, 48 hours and 5 days of infection, respectively; and B, PCR blank with no template DNA. The gel is representative of RAPD PCR genotyping performed on the 31 bacterial samples archived post infection from the nasopharynx or lung (24 BCC0191 WT and 7 BCC0191 Δc3), and all colonies produced identical RAPD PCR profiles.



Supplementary Figure 9. Comparison of the gene organization for the *Burkholderia* cepacin A and *Collimonas* collimomycin BGCs. The predicted core biosynthesis genes are highly conserved between the pathways. The *Collimonas* BGC possesses several additional genes, mostly of hypothetical function, and the regulatory component upstream of the core biosynthetic genes differs between the two species. The sequence comparison image was constructed using the Python application Easyfig[8].



Supplementary Figure 10. Rooted core-gene phylogeny of 64 *B. ambifaria* strains. The phylogenetic tree was constructed based on 1594 core genes identified and aligned using the software Roary[9]. The *B. ambifaria* root was determined by including the outgroup species, *Burkholderia vietnamiensis* G4. FastTree[10] was used to construct the approximate-maximum-likelihood phylogeny using the generalised time reversible substitution model. The evolutionary distance scale bar represents the number of base substitutions per site.

Supplementary Table 1. *B. ambifaria* strains and genomes used during this study.

Study Strain Name	Alternative Name	Source details	Accession No. or Bioproject	Reference
BCC0118	CEP0617; LMG P-24636	CF (Sputum); USA	ERS784989	/
BCC0191	HI 2345 (J82); ATCC 51993; ARS BcB	ENV (Soil); USA	ERS784799	[11]
BCC0192	Ral-3; R-8863; HI2347; FC627	ENV (Corn rhizosphere, biocontrol strain); USA	ERS785047	[12]
BCC0197	ATCC 51671; LMG 19465; FC661; R-9945; B37w	ENV (Leaves of *Sesbania exaltata*, biocontrol strain)	ERS785076	[13]
BCC0200	Formally *B.cepacia* gv I (BC-B)	ENV; USA	ERS785045	[11]
BCC0203	BCF/HG1-A; LMG-P 24640	ENV; USA	ERS782625	[14]
BCC0207	AMMD (LMG 19182^T)	ENV (Pea rhizosphere); USA	PRJNA13490	[12]
BCC0250	CEP0958;LMG P-24637; R-9927	CF (Sputum); Australia	ERS784819	/
BCC0267	LMG 19467; CEP0996; R-9935	CF (Sputum); Australia	ERS784835	[12]
BCC0284	ATCC 53267; LMG 17829; CEP0102; C2965	ENV (Corn roots); USA	ERS1328916	[12]
BCC0316	M54, HI 2347, R-5142	ENV (Soil); USA	ERS784850	[12]
BCC0338	ATCC 53266 LMG 17828; FC662	ENV (Corn roots); USA	ERS1371637	[12]
BCC0399	CEP1054	CF; USA	ERS784860	/
BCC0410	MVP/C1 64	ENV (Maize); Italy	ERS784866	[15]
BCC0423	MCI 4	ENV (Maize); Italy	ERS1336067	[16]
BCC0477	AU0216	CF (Sputum); USA	ERS784882	/
BCC0478	AU1366	CF (Sputum); USA	ERS784897	/
BCC0480	HI-2427	ENV (Soil); USA	ERS784913	/
BCC1041	MVP-C2-51	ENV (Maize); Italy	ERS784930	[15]
BCC1048	MVP-C2-69	ENV (Maize); Italy	ERS784886	[15]
BCC1052	MCII-68	ENV (Maize); Italy	ERS1371632	[17]
BCC1062	MDII-130riz	ENV (Maize); Italy	ERS1328829	[18]
BCC1065	MDIII-B-388	ENV (Maize); Italy	ERS784959	[18]
BCC1066	MDIII-B-399	ENV (Maize); Italy	ERS784800	[18]
BCC1072	MDIII-P-170	ENV (Maize); Italy	ERS1328913	[18]
BCC1080	MDIII-T-2	ENV (Maize); Italy	ERS1371635	[18]
BCC1083	MDIII-T-50	ENV (Maize); Italy	ERS1328835	[18]
BCC1086	MDIII-T-401(s)	ENV (Maize); Italy	ERS1328827	[18]
BCC1088	MDIII-T-474(s)	ENV (Maize); Italy	ERS1371633	[18]
BCC1090	MVP-C1-40	ENV (Maize); Italy	ERS1328833	[15]
BCC1092	MVP-C1-53	ENV (Maize); Italy	ERS784808	[15]
BCC1093	MVP-C1-55	ENV (Maize); Italy	ERS784821	[15]
BCC1095	MVP-C1-80	ENV (Maize); Italy	ERS784837	[15]
BCC1098	MVP-C1-95	ENV (Maize); Italy	ERS784852	[15]
BCC1100	MVP-C2-25	ENV (Maize); Italy	ERS784868	[15]
BCC1103	MVP-C2-44	ENV (Maize); Italy	ERS784884	[15]
BCC1105	MVP-C2-73	ENV (Maize); Italy	ERS1371636	[15]
BCC1107	MVP-C2-79	ENV (Maize); Italy	ERS784899	[15]
BCC1212	MC40-6	ENV (Rhizosphere); USA	PRJNA17411	/
BCC1213	MC80-27	ENV; USA	ERS1328957	/
BCC1214	MA80-5	ENV; USA	ERS784915	/
BCC1216	MW20-13	ENV; USA	ERS1328918	/
BCC1218	MW80-16	ENV; USA	ERS784932	/
BCC1220	MS5-3	ENV; USA	ERS1328837	/
BCC1223	MS80-4	ENV; USA	ERS784947	/
BCC1224	KS0-1	ENV (Maize); USA	ERS1328836	[19]
BCC1228	KA20-1	ENV (Maize); USA	ERS1328832	[19]
BCC1229	KA5-1	ENV (Maize); USA	ERS1371639	[19]
BCC1233	KC0-24	ENV (Maize); USA	ERS1328917	[19]
BCC1236	KC5-54	ENV (Maize); USA	ERS1328839	[19]
BCC1237	KC10-16	ENV (Maize); USA	ERS1371640	[19]
BCC1240	KC311-11	ENV (Maize); USA	ERS1328914	[19]
BCC1241	KC311-6	ENV (Maize); USA	ERS784961	[19]
BCC1246	KC20-40	ENV (Maize); USA	ERS1328834	[19]
BCC1248	KW0-1; LMG-P 24641	ENV (Maize); USA	ERS784801	[19]
BCC1249	KW0-5	ENV (Maize); USA	ERS1371634	[19]
BCC1252	KW10-1	ENV (Maize); USA	ERS784809	[19]
BCC1256	KW420-19	ENV (Maize); USA	ERS784823	[19]
BCC1258	KW318-1	ENV (Maize); USA	ERS1371641	[19]
BCC1259	KW20-2	ENV (Maize); USA	ERS784838	[19]
BCC1265	MC40-7	ENV; USA	ERS784854	/
BCC1270	KC20-17	ENV (Maize); USA	ERS784870	[19]
IOP40-10	/	ENV (Prairie grass rhizosphere)	PRJNA20669	/
MEX-5	/	ENV (Teosinte plants (*Zea perennis*)	PRJNA20667	/
BCC0191 ::*ccnJ*	/	/	/	This study
BCC1252 ::*ccnJ*	/	/	/	This study
BCC1241 ::*ccnJ*	/	/	/	This study
BCC0477 ::*ccnJ*	/	/	/	This study
BCC1259 ::*ccnJ*	/	/	/	This study
BCC1218 ::*ccnJ*	/	/	/	This study
BCC0191Δc3	/	/	/	This study
BCC0191 ::*ccnJ*Δc3	/	/	/	This study

Supplementary Table 2. *B. ambifaria* genomic statistics.

Strain	Total Contigs (Mbp)	Mapped Contigs (Mbp)	Total Contig No.	Contigs <1000 bp	N50 (bp)
BCC0118	7.50	7.43	103	45	273879
BCC0191	7.58	7.55	97	57	302202
BCC0192	7.40	7.32	100	48	277504
BCC0197	7.38	7.36	82	33	386867
BCC0200	7.63	7.55	124	64	292321
BCC0203	7.93	7.53	4	4	2669373
BCC0207	7.53	7.48	4	4	2646969
BCC0250	7.36	7.33	88	32	396147
BCC0267	7.36	7.33	74	27	784052
BCC0284	7.47	7.38	88	34	423851
BCC0316	7.64	7.56	131	62	292321
BCC0338	7.45	7.42	61	28	514723
BCC0399	7.40	7.38	86	38	407605
BCC0410	7.38	7.36	44	18	849830
BCC0423	7.47	7.30	100	36	382759
BCC0477	7.81	7.43	118	54	255598
BCC0478	7.24	7.21	89	25	601161
BCC0480	7.84	7.52	87	24	570779
BCC1041	7.51	7.42	88	38	469567
BCC1048	7.51	7.43	76	35	469249
BCC1052	7.33	7.30	77	34	397259
BCC1062	7.45	7.36	90	37	382455
BCC1065	7.31	7.28	81	32	605534
BCC1066	7.32	7.29	112	37	381533
BCC1072	7.44	7.36	93	40	397405
BCC1080	7.29	7.26	85	35	383339
BCC1083	7.30	7.26	95	35	373328
BCC1086	7.60	7.40	302	49	395516
BCC1088	7.48	7.40	70	38	397405
BCC1090	7.33	7.29	114	33	340920
BCC1092	7.38	7.36	57	20	849686
BCC1093	7.40	7.38	91	37	427669
BCC1095	7.40	7.38	79	34	407577
BCC1098	7.38	7.37	48	15	849506
BCC1100	7.40	7.37	91	37	407581
BCC1103	7.40	7.37	77	38	427387
BCC1105	6.30	6.13	50	15	1592784
BCC1107	7.40	7.37	83	39	374059
BCC1213	7.47	7.42	107	40	425516
BCC1214	7.61	7.14	169	79	270842
BCC1216	7.46	7.33	166	73	234327
BCC1218	7.46	7.42	109	51	348377
BCC1220	7.64	7.14	188	74	393412
BCC1223	7.61	7.14	157	80	267383
BCC1224	7.27	7.19	152	53	360040
BCC1228	7.38	7.35	82	33	558864
BCC1229	7.51	7.48	100	51	298531
BCC1233	7.96	7.42	126	75	390415
BCC1236	7.63	7.58	123	47	358767
BCC1237	7.34	7.31	91	32	420767
BCC1240	7.40	7.37	63	29	503900
BCC1241	7.68	7.44	138	61	288780
BCC1246	7.48	7.44	72	32	450462
BCC1248	8.03	7.50	273	132	241997
BCC1249	7.58	7.55	71	27	447443
BCC1252	7.42	7.38	107	60	254705
BCC1256	7.54	7.47	104	49	281063
BCC1258	7.60	7.52	110	57	272850
BCC1259	7.63	7.30	162	75	431727
BCC1265	7.62	7.29	136	50	335885
BCC1270	7.60	7.53	115	51	315858
IOP40-10	7.69	6.88	629	571	24355
MC40-6	7.64	7.34	4	4	2769414
MEX-5	7.86	6.58	706	634	20649

Supplementary Table 3. De-replicated secondary metabolite gene clusters of *B. ambifaria*.

Cluster Type (antiSMASH prediction)	Prevalence in *B. ambifaria* (out of 64 genomes)	Average length (kbp)[a,b]
Replicon c1		
Terpene 1	65[c]	20.9
NRPS 1	64	54.5
Arylpolyene 1	64	41.2
NRPS 2	44	46.8
PKS 1	28	47.6
Lantipeptide 1	10	27.1
Butyrolactone 1	1	11.0
Terpene 2	1	21.0
Replicon c2		
Homoserine lactone 1	64	20.6
Phosphonate	64	41.7
Terpene 3	64	21.1
Other 1	64	41.1
Terpene 4	64	24.0
Bacteriocin 1	55	13.1
Arylpolyene 2	53	44.9
Ectoine 1	53	10.4
Ectoine 2	38	10.4
Homoserine lactone (cepacin-associated) 2	22	20.7
Other 2	20	43.0
Butyrolactone-otherKS	16	32.5
Phenazine	13	20.4
PKS 2	7	44.9
NRPS 3	2	52.8
Bacteriocin 2	1	25.9
Arylpolyene 3	1	41.1
Replicon c3		
Homoserine lactone 3	63	20.6
Bacteriocin 3	63	10.8
Terpene 5	63	22.0
NRPS-T1PKS 1	54[d]	85.5 or 117.6[e]
PKS 3	40	43.9 and 63.9[f]
Other 3	34	27.0
Butyrolactone 2	27	10.7
Other 4	24	43.8
NRPS-T1PKS 2	17	60.9
Trans-AT-PKS	6	91.2
Unknown genomic location		
NRPS 4	1	26.8
Homoserine lactone 4	1	20.1
NRPS-T1PKS-*trans*-AT-PKS	1	69.8

Footnotes:

[a] *B. ambifaria* strains MEX-5 and IOP40-10 were excluded from several cluster length averages due to the fragmented nature of clusters caused by lower quality genome assemblies. Other strains were excluded if the clusters were manually split from the antiSMASH predicted cluster.

[b] Average length was calculated from predicted antiSMASH clusters or extracted sequences identified using BLAST; partial clusters were excluded.

[c] *B. ambifaria* strain BCC1249 encoded a duplicate terpene 1 cluster.

[d] *B. ambifaria* strains MEX-5, IOP40-10 and BCC1065 either possess low quality genomes preventing the reconstruction of the burkholdine NRPS-T1PKS pathway or encode a partial cluster.

[e] The published NRPS-T1PKS pathway (85.5 kbp) was encoded by 36 *B. ambifaria* strains, while the larger 117.6 kbp pathway was encoded by 15 *B. ambifaria* strains and consists of the published pathway with additional biosynthetic genes.

[f] The 43.9 kbp version of the pathway was encoded by 26 *B. ambifaria* strains, while the 63.9 kbp version was encoded by 14 *B. ambifaria* strains.

Supplementary Table 4. Known antimicrobial compounds produced by *B. ambifaria*.

Compound	Prevalence in *B. ambifaria*	Replicon	Bioactivity activity
Pyrrolnitrin	100%	C2	Anti-fungal
Burkholdine	84%	C3	Anti-fungal
AFC-BC11	53%	C3	Anti-fungal
Hydroxyquinolines	38%	C3	Signaling molecule/Anti-fungal
Cepacin A[a]	34%	C2	Anti-oomycetal/Anti-Gram-positive
Bactobolins	27%	C3	Anti-Gram-negative/Anti-Gram-positive
Phenazine	20%	C2	Anti-fungal
Enacyloxin IIa	9%	C3	Anti-Gram-negative

Footnotes:
[a] Cepacin A biosynthetic gene cluster was identified during this study.

Supplementary Table 5. Antimicrobial susceptibility organisms: host and disease phenotypes[1–4] and incubation temperatures.

Organism	Source/ID Number	Incubation Temperature (°C)	Host Range	Disease
Pathogenic bacteria and fungi				
Rhizobium radiobacter	LMG 187	30	Broad	Crown gall disease
Pseudomonas savastanoi pv. *phaseolicola*	LMG 2245	30	Common bean	Bean halo blight
Dickeya solani	LMG 25993	30	European potato	Potato tuber rot
Pectobacterium carotovorum	LMG 2464	30	Several crop species	Soft rot disease
Pseudomonas syringae pv. *tomato*	LMG 5093	30	Tomato	Bacterial speck
Xanthomonas campestris pv. *campestris*	8004	30	Cultivated Brassicaceae	Black rot
Pseudomonas syringae pv. *syringae*	LMG 187	30	Common bean	Bacterial brown spot
Burkholderia multivorans	ATCC 17616	30	Human (immunocompromised)	Opportunistic (e.g. CF lung infection)
Staphylococcus aureus	NCTC 12981	37	Human	Multiple (cutaneous, systemic etc.)
Enterococcus faecalis	ATCC 51299	37	Human	Multiple (uninary, systemic etc.)
Candida albicans	SC 5314	37	Human	Candidiasis
Fusarium solani var. redolens (Wollenweber)	MUCL 14241	22	Multiple crop species	Root/foot rot, wilt
Alternaria alternata (Fries:Fries) von Keissler	MUCL 36	22	Several (pathovar specific)	Black-spot, stem canker
Other reference organisms				
Bacillus subtilis	ATCC 23857	30	N/A	N/A

Supplementary Table 6. Minimum Inhibitory Concentration (MIC) of enacyloxin IIa against plant and animal pathogens.

Organism	MIC (µg/ml)
Rhizobium radiobacter	3.2
Pseudomonas savastanoi pv. *phaseolicola*	50.0
Dickeya solani	12.5
Pectobacterium carotovorum	6.3
Pseudomonas syringae pv. *tomato*	50.0
Xanthamonas campestris pv. *campestris*	50.0
Pseudomonas syringae pv. *syringae*	6.3
Burkholderia multivorans	6.3

Footnotes:
MICs were measured in a microbroth dilution assay using iso-sensitest broth or TSB broth with doubling-dilutions of enacyloxin IIa between 100 and 0.098 µg/ml. Bacteria were grown for 18-24 hours and the optical density at 600 nm was measured. The MIC was determined by calculating the enacyloxin IIa concentration required to produce an 80% knockdown in optical density compared to the organism grown in the absence of the antibiotic; n = 3 replicates for MIC analysis.

Supplementary Table 7. The rhizocompetence of *B. ambifaria* BCC0191 and derived mutants.

Replicate and Statistics	Colony forming units per gram of root [a]		
	WT	**::*ccnJ***	**Δc3**
Replicate 1	$3.1x10^7$	$1.0x10^7$	$9.0x10^6$
Replicate 2	$1.7x10^7$	$7.9x10^6$	$1.1x10^7$
Replicate 3	$3.6x10^7$	$2.1x10^8$	$5.4x10^6$
Mean[b]	$2.8x10^7$	$7.8x10^7$	$8.5x10^6$

Footnote:

[a] The total viable counts of each replicate (n = 3 plants per condition) from BCC0191 wild-type and mutant derivatives cultured from 14 day colonised root segments and adjusted to 1 g fresh weight of root is shown. No *B. ambifaria* growth was detected on the control root segments.

[b] Significant difference (P = 0.027, df = 4, t = -4.40, 95% confidence interval), between WT and Δc3 as determined by two-sided (unpaired) t-test assuming normally distributed data (Shapiro-Wilk test) and equal variances (Bartlett test). Failing these assumptions, the non-parametric Wilcoxon rank sum test was performed to determine significance.

Supplementary Table 8. Proteins with similarity to those encoded by the cepacin A biosynthetic gene cluster.

Gene Name	Length bp/aa	Similar Proteins (excluding *B. ambifaria*)	% aa Sequence Identity	Proposed Function
ccnA	714/237	LuxR family transcriptional regulator [Burkholderia vietnamiensis] LuxR family transcriptional regulator [Burkholderia contaminans] LuxR family transcriptional regulator [Burkholderia sp. USM B20]	82% 83% 82%	LuxR family transcriptional regulator
ccnB	384/127	Hypothetical Protein [Burkholderia sp. LA-2-3-30-S1-D2] Hypothetical Protein [Burkholderia sp. RF2-non_BP3] Hypothetical Protein [Burkholderia sp. RF4-BP95]	90% 87% 88%	Hypothetical
ccnC	654/217	GNAT family N-acetyltransferase [Burkholderia sp. LA-2-3-30-S1-D2] GNAT family N-acetyltransferase [Burkholderia sp. RF4-BP95] GNAT family N-acetyltransferase [Burkholderia contaminans]	88% 87% 87%	LuxI homoserine lactone synthase
ccnD	1293/430	beta-ketoacyl-ACP synthase II [Burkholderia sp. LA-2-3-30-S1-D2] beta-ketoacyl-ACP synthase II [Burkholderia sp. RF4-BP95] beta-ketoacyl-ACP synthase II [Burkholderia sp. RF2-non_BP3]	95% 94% 94%	Beta-ketoacyl synthase
ccnE	1386/461	NAD(P)/FAD-dependent oxidoreductase [Burkholderia contaminans] NAD(P)/FAD-dependent oxidoreductase [Burkholderia vietnamiensis] NAD(P)/FAD-dependent oxidoreductase [Burkholderia vietnamiensis]	88% 85% 84%	Flavin-dependent monooxygenase
ccnF	1083/360	fatty acid desaturase [Burkholderia sp. LA-2-3-30-S1-D2] fatty acid desaturase [Burkholderia sp. RF4-BP95] fatty acid desaturase [Burkholderia vietnamiensis]	92% 92% 89%	Fatty acid desaturase (alkane hydroxylase)
ccnG	1005/334	aromatic ring-hydroxylating dioxygenase subunit alpha [Burkholderia contaminans] aromatic ring-hydroxylating dioxygenase subunit alpha [Burkholderia sp. LA-2-3-30-S1-D2] Rieske (2Fe-2S) domain protein [Burkholderia vietnamiensis G4]	95% 94% 94%	Rieske non-heme iron-dependent oxygenase
ccnH	876/291	hypothetical protein [Burkholderia contaminans] hypothetical protein [Burkholderia sp. LA-2-3-30-S1-D2] hypothetical protein [Burkholderia sp. RF2-non_BP3]	86% 82% 81%	Hypothetical
ccnI	1419/472	MFS transporter [Burkholderia sp. LA-2-3-30-S1-D2] MFS transporter [Burkholderia sp. RF4-BP95] MFS transporter [Burkholderia sp. RF2-non_BP3]	90% 90% 90%	MFS transporter
ccnJ	1797/598	AMP-dependent synthetase [Burkholderia contaminans] AMP-dependent synthetase [Burkholderia sp. RF4-BP95] AMP-dependent synthetase [Burkholderia vietnamiensis]	88% 86% 85%	Fatty acyl-AMP ligase
ccnK	960/319	acyl-CoA desaturase [Burkholderia contaminans] acyl-CoA desaturase [Burkholderia sp. RF4-BP95] acyl-CoA desaturase [Burkholderia sp. LA-2-3-30-S1-D2]	92% 93% 92%	Fatty acid desaturase
ccnL	984/327	acyl-CoA desaturase [Burkholderia contaminans] acyl-CoA desaturase [Burkholderia sp. RF4-BP95] acyl-CoA desaturase [Burkholderia sp. RF2-non_BP3]	96% 95% 94%	Fatty acid desaturase
ccnM	321/106	polyketide synthase [Burkholderia sp. RF2-non_BP3] polyketide synthase [Burkholderia sp. LA-2-3-30-S1-D2] polyketide synthase [Burkholderia vietnamiensis]	96% 95% 95%	Acyl carrier protein
ccnn	1098/365	fatty acid desaturase [Burkholderia contaminans] fatty acid desaturase [Burkholderia stagnalis] fatty acid desaturase [Burkholderia stagnalis]	94% 92% 91%	Fatty acid desaturase
ccnO	924/307	delta-12-desaturase [Burkholderia sp. LA-2-3-30-S1-D2] delta-12-desaturase [Burkholderia vietnamiensis] delta-12-desaturase [Burkholderia contaminans]	87% 86% 87%	Thioesterase
ccnP	186/61	rubredoxin [Burkholderia vietnamiensis] rubredoxin [Burkholderia contaminans] rubredoxin [Burkholderia stagnalis]	95% 93% 92%	Rubredoxin

Supplementary Table 9. Primers used during this study.

Primers[a]	Final Product Size	Primer Use	Source
Fwd: 5'- GCG TCT AGA GAC GTG ATC ATT GCC GGA AA -3' Rev: 5'- GCG GAA TTC TTG CCC GAT ACA TAG AGC GT -3'	707	Amplify product from fatty AMP ligase-encoding gene, *ccnJ*	This study
Fwd: 5'- TTA YTT TTG YGC CGC TAC MG -3' Rev: 5'- CCM GAG CAG CTY TAT ACG AT -3'	582	Screen for presence of c3 replicon in *B. ambifaria* BCC0191Δc3 (degenerate for *B. ambifaria*)	This study
Fwd: 5'- AAG AAA TCT GCT GCC GCT TG -3' Rev: 5'- CAC TTC GCT GTA CCT CAA GC -3'	608	Screen for presence of pMinic3 in *B. ambifaria* BCC0191Δc3	This study
5'- AGC GGG CCA A -3'	Variable	Random amplified polymorphic DNA genotyping	Mahenthiralingam *et al.*[7]

Footnotes:
[a] Restriction sites in primer sequences are underlined.

Supplementary Notes

Genome sequencing quality control and assembly. Illumina reads were trimmed using the wrapper script Trim Galore v0.4.2[20], which utilises Cutadapt v1.12[21], and their quality assessed using FastQC v0.10.1[22]. FLaSH v1.2.11[23] was used to merge overlapping short read pairs to improve contiguity of assembled genomes. The resulting overlapped and paired-end reads output from FLASH were assembled into contigs using the assembler SPAdes v3.9.1[24]. Misassembled contigs were identified and corrected with Pilon v1.21[25]. Contig sequences representing contaminating DNA were identified with Kraken v0.10.5-beta[26] using the Minikraken database and removed prior to genomic analyses. Genome sequence quality assessment and statistics were calculated using QUAST v4.4[27].

B. ambifaria **genomics and** ***in silico*** **definition of specialized metabolite biosynthetic gene clusters.** AntiSMASH analysis[28] and BLAST[29] searches for known specialized metabolite biosynthetic gene clusters (BGCs) detected a total of 1,272 BGCs across 64 *B. ambifaria* strains, defining a mean of 20 pathways per genome. Eighteen classes of metabolites were identified as well as one distinct BGC that was not recognised as encoding a previously reported metabolite class by antiSMASH, but was recognised as having the potential to direct the production of a specialized metabolite (there were 20 examples of this BGC in the genome set; see Figure 2). A combination of Kmer-matching and gene topology comparisons enabled de-replication of the 1,272 BGCs to 38 distinct putative BGCs (Figure 2; Supplementary Table 3). Of the 38 distinct BGCs, three singleton pathways were detected in contigs that had no significant homology to the reference sequences, and could not be incorporated into the assembled replicons c1, c2 and c3 (Figure 2; Supplementary Table 3).

The specialized metabolite encoding capacity of the replicons relative to their size (density) varied significantly (Supplementary Figure 1). The largest replicon (c1) possessed the lowest specialized metabolite BGC density, whereas the smallest replicon (c3) possessed the largest density. The metabolite BGC density of the replicons was not static across the *B. ambifaria* dataset, with replicon c3 varying in BGC density from less than 10% to above 30% (mean 19.4%; Supplementary Figure 1). Replicons c1 and c2 displayed a lower mean BGC density and density variance compared to c3. *B. ambifaria* strains encoded 3-6 BGCs on replicon c1, with eight distinct BGCs identified on the replicon. Replicon c2 encoded 8-11 BGCs with 17 distinct BGCs; and c3 encoded 4-9 BGCs with eleven distinct BGCs detected.

Of the 38 distinct BGCs, three singleton pathways were detected in contigs that had no significant homology to the reference sequences, and could not be incorporated into the assembled replicons c1, c2 and c3 (Figure 2; Supplementary Table 3). One of BGC encoded a hybrid non-ribosomal peptide synthetase *trans*-acyltransferase polyketide synthase (NRPS-*trans*-AT PKS), and showed 92% sequence similarity to the malleilactone biosynthetic gene cluster[30,31]. The two remaining BGCs with unknown genomic locations were an uncharacterised NRPS in strain MEX-5, and a LuxRI system in strain IOP40-10. The most frequently detected specialized metabolite classes were terpenes (5/38) and NRPS (4/38). Eleven of the distinct BGCs were encoded by all 64 *B. ambifaria* strains, and included four terpene synthase BGCs and two LuxRI systems (Figure 2). Eight clusters were detected in less than 5% of the *B. ambifaria* strains examined (<3 strains), three of which were type 1 modular PKS gene clusters (Figure 2).

Analysis of QS-regulated BGCs in ***B. ambifaria.*** The availability of the extensive *B. ambifaria* genome and BGC datasets enabled interrogation of QS regulatory genes with a focus on LuxR-encoding genes as the key BGC regulators. We detected 356 *luxR* homologues across the 64 *B. ambifaria* strains, representing 14 distinct protein phylogenetic clades (Figure 3). These clades included the following *B. ambifaria* LuxRI quorum sensing (QS) systems: the *bafRI* system[32] (63 of 64 strains), the *cepR2I2*

system[33] (61 of 64 strains), one functionally uncharacterised system (22 of 64 strains), and a QS system present only in *B. ambifaria* IOP40-10. Six LuxR clades were associated with BGCs encoding the following compounds or compound classes (Figure 3): ectoine, lantipeptide, butyrolactone, enacyloxin IIa, bactobolins and a putative BGC that was identified as directing cepacin biosynthesis (see Figure 4). The remaining LuxR clades flanked membrane transporter genes, a type 3 secretion system and four clades were associated with genes of unknown collective functions (Figure 3).

Phenotypic analysis of *B. ambifaria* BCC0191, cepacin and third replicon mutants. Antimicrobial activity against the panel of plant and animal pathogenic bacteria and fungi, as well as further reference strains, was examined. The loss of cepacin A production in BCC0191::*ccnJ* resulted in loss of anti-Gram-positive activity against *Staphylococcus aureus*, *Enterococcus faecalis* and *Bacillus subtilis* (Supplementary Figure 6), in addition to the loss of *Pythium* inhibition and the weak activity against certain Gram-negative bacteria (Figure 4b; Supplementary Figure 4). Deletion of the third replicon resulted in loss of antagonism against the fungal species *Candida albicans*, *Fusarium solani* and *Alternaria alternata*, but enhanced anti-Gram positive activity which correlated with a 2-fold increase in cepacin production seen in the BCC0191Δc3 mutant (Supplementary Figure 6). The double mutant, BCC0191::*ccnJ*Δc3, lost all the antimicrobial phenotypes observed for the wild-type *B. ambifaria* BCC0191 (Supplementary Figure 6). The rhizocompetence of *B. ambifaria* BCC0191 WT, BCC0191::*ccnJ* and BCC0191Δc3 mutants was also evaluated to assess whether their effect on biological control of *Pythium* (Figure 5) was a consequence of reduced root colonisation (see Supplementary Methods). After 14 days of growth in the pea biological control model system (not containing *P. ultimum*), the wild-type and BCC0191::*ccnJ* colonised the rhizosphere at equivalent levels ($P = 0.7$; Supplementary Table 7). The BCC0191Δc3 mutant colonised the pea rhizosphere at a rate significantly lower ($P = 0.027$, df = 4, t = -4.40), but within 0.5 log of the mean wild-type level of 2.8×10^7 cfu/g root (Supplementary Table 7).

Supplementary Discussion

Pan-genomics and extensive specialized metabolite diversity. Genomic analysis across 64 *B. ambifaria* genomes revealed substantial diversity in gene content and predicted specialized metabolite BGCs; branching into five defined clades. Eleven of the 38 BGCs were encoded by all studied *B. ambifaria* strains, the remaining BGCs constitute the accessory specialized metabolite potential of *B. ambifaria* (see Figure 2). The large core and accessory metabolite potential in *B. ambifaria* compliments previous studies of the wider *Burkholderia* genus[34]. Similar in-depth, species-focussed BGC distribution analyses were conducted in three *Salinospora* species with previously unknown metabolite potential, but these were restricted to NRPS, PKS and fatty acid synthase pathways[35]. A more limited analysis (6 strains) was also conducted with *Streptomyces albus* genomes [36]. Pan-genomic analysis of nine strains has been performed on the biocontrol species *Pseudomonas putida* [37], but a detailed correlation to their specialized metabolite biosynthetic capacity and protective properties, such as that reported in this study, was lacking.

Multifaceted analysis of previously characterised biocontrol strains. We analysed the *in vitro* antimicrobial activity, specialized metabolite BGCs, and metabolite profiles of ten strains spanning the *B. ambifaria* core-gene phylogeny, including seven of the eight previously characterised biological control strains (Supplementary Table 1): AMMD (BCC0207), BC-F (BCC0203), J82 (BCC0191), M54 (BCC0316), Ral-3 (BCC0192), ATCC 53267 (BCC0284) and ATCC 53266 (BCC0338). Most of these biocontrol strains were initially characterised by agricultural and biotechnology companies as potent agents capable of suppressing multiple plant pathogens. Variation was observed in the antimicrobial BGC content of the ten strains. The presence of multiple BGCs encoding broad spectrum antimicrobials suggests a “built-in” redundancy to their biocontrol ability to suppress plant pathogens. Despite encoding antimicrobial BGCs,

there were several examples of non-producing BGCs under the experimental conditions used. The absence of these metabolites in encoding strains may be due to multiple factors such as the presence of BGC or regulatory mutations within these strains, the need for specific growth conditions to prime biosynthesis, or that BGC expression is activated by complex interactions such as contact with a competing organism or inter-kingdom signalling from the host plant.

Effect of third replicon deletion on the pathogenicity of *B. ambifaria* BCC0191. For *B. ambifaria* strains such as AMMD, deletion of the third genomic replicon rendered the derived mutant, AMMDΔc3, attenuated in multiple virulence models and resulted in loss of anti-fungal activity[38], but retained the ability to colonise the rhizosphere[39]. Our *B. ambifaria* BCC0191Δc3 mutant exhibited a non-significant reduction in biocontrol efficacy against *P. ultimum* (10^5: $P = 0.22$, $t = -1.44$, $df = 4$; 10^6: $P = 0.22$, $t = -1.46$, $df = 4$; 10^7: $P = 0.16$, , $t = -1.73$, $df = 4$; Figure 5b), but exhibited a significantly reduced rhizocompetence, with the number of viable bacteria per g of root an order of magnitude below that of the wild-type ($P = 0.027$, $t = -3.40$, $df = 4$; Supplementary Table 7). In contrast to AMMDΔc3[38], *B. ambifaria* BCC0191Δc3 retained virulence in a *Galleria* wax-moth larvae virulence assay (Supplementary Figure 7a), suggesting in strain BCC0191 *Galleria* virulence is mediated by factors encoded on the first and second replicons. However, corroborating the reduced pathogenicity observed for c3 replicon mutants in vertebrate models[38], *B. ambifaria* BCC0191Δc3 was not able to persist within the lungs of infected mice (Supplementary Figure 7c). At the equivalent infective dose (2 x10^6 bacteria), highly virulent respiratory pathogens such as *Pseudomonas aeruginosa* persist in the same murine infection model at log-fold higher densities in both nasopharynx and lung[40,41]. In addition, no *B. ambifaria* was detected within the spleens of infected mice suggesting that the capability of this species for invasive infection, observed during fatal “cepacia syndrome” CF infection[42], is low.

Supplementary References

1. Howden, A. J. M., Rico, A., Mentlak, T., Miguet, L. & Preston, G. M. *Pseudomonas syringae* pv. *syringae* B728a hydrolyses indole-3-acetonitrile to the plant hormone indole-3-acetic acid. *Mol. Plant Pathol.* **10,** 857–865 (2009).
2. Mansfield, J. *et al.* Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* **13,** 614–629 (2012).
3. Coleman, J. J. The *F usarium solani* species complex: ubiquitous pathogens of agricultural importance. *Mol. Plant Pathol.* **17,** 146–158 (2016).
4. Tsuge, T. *et al.* Host-selective toxins produced by the plant pathogenic fungus *Alternaria alternata*. *FEMS Microbiol. Rev.* **37,** 44–66 (2013).
5. Spitzer, M., Wildenhain, J., Rappsilber, J. & Tyers, M. BoxPlotR: a web tool for generation of box plots. *Nat. Methods* **11,** 121–2 (2014).
6. Dalton, J. P., Uy, B., Swift, S. & Wiles, S. A novel restraint device for injection of *Galleria mellonella* larvae that minimizes the risk of accidental operator needle stick injury. *Front. Cell. Infect. Microbiol.* **7,** 99 (2017).
7. Mahenthiralingam, E., Campbell, M. E., Henry, D. A. & Speert, D. P. Epidemiology of *Burkholderia cepacia* infection in patients with cystic fibrosis: analysis by randomly amplified polymorphic DNA fingerprinting. *J. Clin. Microbiol.* **34,** 2914–20 (1996).
8. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27,** 1009–10 (2011).
9. Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31,** 3691–3693 (2015).
10. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* **5,** e9490 (2010).
11. Mao, W., Lewis, J. A., Hebbar, P. K. & Lumsden, R. D. Seed Treatment with a Fungal or a Bacterial Antagonist for Reducing Corn Damping-off Caused by Species of *Pythium* and *Fusarium*. *Plant Dis.* **81,** 450–454 (1997).
12. Coenye, T. *et al. Burkholderia ambifaria* sp. nov., a novel member of the Burkholderia cepacia complex including biocontrol and cystic fibrosis-related isolates. *Int. J. Syst. Evol. Microbiol.* **51,** 1481–1490 (2001).
13. Schisler, D. A., Howard, K. M. & Bothast, R. J. Enhancement of disease caused by *Colletotrichum truncatum* in *Sesbania exaltata* by coinoculating with epiphytic bacteria. *Biol. Control* **1,** 261–268 (1991).
14. Mao, W., Lewis, J. A., Lumsden, R. D. & Hebbar, K. P. Biocontrol of selected soilborne diseases of tomato and pepper plants. *Crop Prot.* **17,** 535–542 (1998).
15. Dalmastri, Chiarini, Cantale, Bevivino & Tabacchioni. Soil Type and Maize Cultivar Affect the Genetic Diversity of Maize Root-Associated *Burkholderia cepacia* Populations. *Microb. Ecol.* **38,** 273–284 (1999).
16. Bevivino, A. *et al.* Characterization of a free-living maize-rhizosphere population of *Burkholderia cepacia*: effect of seed treatment on disease suppression and growth promotion of maize. *FEMS Microbiol. Ecol.* **27,** 225–237 (1998).
17. Di Cello, F. *et al.* Biodiversity of a *Burkholderia cepacia* population isolated from the maize rhizosphere at different plant growth stages. *Appl. Environ. Microbiol.* **63,** 4485–93 (1997).
18. Pirone, L., Chiarini, L., Dalmastri, C., Bevivino, A. & Tabacchioni, S. Detection of cultured and uncultured *Burkholderia cepacia* complex bacteria naturally occurring in the maize rhizosphere. *Environ. Microbiol.* **7,** 1734–1742 (2005).
19. Ramette, A. & Tiedje, J. M. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc. Natl. Acad. Sci.* **104,** 2761–2766 (2007).
20. Krueger, F. Trim Galore! *Babraham Bioinformatics* (2016). Available at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. (Accessed: 7th April 2017)
21. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10 (2011).
22. Andrews, S. (Babraham B. FastQC A Quality Control tool for High Throughput Sequence Data. *Babraham Bioinformatics* (2009).
23. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27,** 2957–63 (2011).
24. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–77 (2012).
25. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9,** e112963 (2014).
26. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15,** (2014).
27. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1075 (2013).
28. Weber, T. *et al.* antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43,** W237-43 (2015).
29. Morgulis, A. *et al.* BLAST+: architecture and applications. *Bioinformatics* **24,** 1757–1764 (2008).
30. Biggins, J. B., Ternei, M. A. & Brady, S. F. Malleilactone, a polyketide synthase-derived virulence factor encoded by the cryptic secondary metabolome of *Burkholderia pseudomallei* group pathogens. *J. Am. Chem. Soc.* **134,** 13192–13195 (2012).
31. Franke, J., Ishida, K. & Hertweck, C. Genomics-driven discovery of burkholderic acid, a noncanonical, cryptic polyketide from human pathogenic *Burkholderia* species. *Angew. Chemie Int. Ed.* **51,** 11611–11615 (2012).
32. Aguilar, C., Friscina, A., Devescovi, G., Kojic, M. & Venturi, V. Identification of quorum-sensing-regulated genes of *Burkholderia cepacia*. *J. Bacteriol.* **185,** 6456–62 (2003).
33. Chapalain, A. *et al.* Interplay between 4-Hydroxy-3-Methyl-2-Alkylquinoline and N-Acyl-Homoserine Lactone Signaling in a *Burkholderia cepacia* complex clinincal strain. *Front. Microbiol.* **8,** 1021 (2017).
34. Depoorter, E. *et al. Burkholderia*: an update on taxonomy and biotechnological potential as antibiotic producers. *Appl. Microbiol. Biotechnol.* **100,** 5215–5229 (2016).
35. Udwary, D. W. *et al.* Genome sequencing reveals complex secondary metabolome in the marine actinomycete

Salinispora tropica. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 10376–81 (2007).
36. Seipke, R. F. Strain-Level Diversity of Secondary Metabolism in *Streptomyces albus*. *PLoS One* **10,** e0116457 (2015).
37. Udaondo, Z., Molina, L., Segura, A., Duque, E. & Ramos, J. L. Analysis of the core genome and pangenome of *Pseudomonas putida*. *Environ. Microbiol.* **18,** 3268–3283 (2016).
38. Agnoli, K. *et al.* Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Mol. Microbiol.* **83,** 362–378 (2012).
39. Vidal-Quist, J. C. *et al. Arabidopsis thaliana* and *Pisum sativum* models demonstrate that root colonization is an intrinsic trait of *Burkholderia cepacia* complex bacteria. *Microbiology* **160,** 373–384 (2014).
40. Fothergill, J. L., Neill, D. R., Loman, N., Winstanley, C. & Kadioglu, A. *Pseudomonas aeruginosa* adaptation in the nasopharyngeal reservoir leads to migration and persistence in the lungs. *Nat. Commun.* **5,** 4780 (2014).
41. Bricio-Moreno, L. *et al.* Evolutionary trade-offs associated with loss of PmrB function in host-adapted *Pseudomonas aeruginosa*. *Nat. Commun.* **9,** 2635 (2018).
42. Parke, J. L. & Gurian-Sherman, D. Diversity of the *Burkholderia cepacia* complex and implications for risk assessment of biological control strains. *Annu. Rev. Phytopathol.* **39,** 225–258 (2001).